Bridging the gap between SKOS and TBX

Detlef Reineke, Laurent Romary

Bridging the gap between SKOS and TBX

Detlef Reineke and Laurent Romary

This article provides an in-depth comparison and proposal for mapping between Simple Knowledge Organization System (SKOS) and TermBase eXchange (TBX), two important exchange standards within the knowledge and terminology landscape. The attempt to develop an interface or conversion routine between SKOS and TBX is rooted in a strong demand in the language and knowledge industries for resource leverage and based on the premise that the two formalisms are governed by similar data models, namely the description of concepts (rather than words).

Keywords: Data exchange, knowledge organization, SKOS, terminology, TBX

Introduction

Over the last decades, the representation of specialized vocabulary in digital format has been on the agenda of several communities that have mostly been working independently from one another:

- Initiated by the need to provide indexing terms for cataloguing, the library community made an initial attempt at defining standards for the representation of thesauri [see 1 and 2], seen as a set of specialized lexical items organized within a tree-based representation. Influenced by further developments in conceptual representations [3], this has led to the development of so-called knowledge organization systems [4] and of SKOS as a linked open data model for lexical information [5].
- The translation and technical writing communities have also heavily invested in developing methods and tools to maintain mono- and multilingual terminologies and have thus produced a whole family of standards to facilitate their management [see 6, 7, 8, 9, and 10, among others].

However, there has been hardly any significant effort to build interfaces for exchanging concept-orientated data between the aforementioned communities. To the best of our knowledge, the only exceptions have been an auspicious mapping proposal by Reineke [11] for data exchange between TBX and RDF/XML and an attempt by the W3C group on Best Practices for Multilingual Linked Open Data to map TBX onto OntoLex [see 12 and 13]. However, the results of the latter are quite unsatisfactory, since OntoLex is essentially a semasiological (word-to-sense) model which, by definition, cannot be naturally mapped onto the concept-to-term model of TBX. As a consequence, the mapping mainly consists of inserting TBX components into various unrelated places in the OntoLex format.

The present article compares the two trends described above, focusing on SKOS and TBX, which, in our view, are both concept-orientated applications that will have to interoperate closely in the future. When referring to SKOS, we focus on the vocabulary listed in the SKOS reference [5], section 2 (SKOS Namespace and Vocabulary) and Appendix B (SKOS eXtension for Labels). TBX is tantamount and limited here to the attribute style as defined in ISO 30042 [10]. The terminological data categories and their respective specifications are replicated from the data category repository DatCatInfo [14]. Both the SKOS and the TBX vocabularies discussed in this article are referenced using notations that point to the respective specifications compiled into a single, more reader-friendly resource available at https://hal.inria.fr/hal-01883377v3/document. As for typography, in running text, markup style is used for SKOS elements (<skos:Concept>), terminological data categories are indicated between forward slashes (/definition/).

Overview of data models

In the language industry, two general classes of data models prevail when providing lexical or terminological information. Semasiological models (or word-to-sense models) are organized with the word as the entry point to a lexical representation, from which further descriptions are derived. This is the model of traditional human-readable dictionaries, where the various senses of a lexical item are grouped

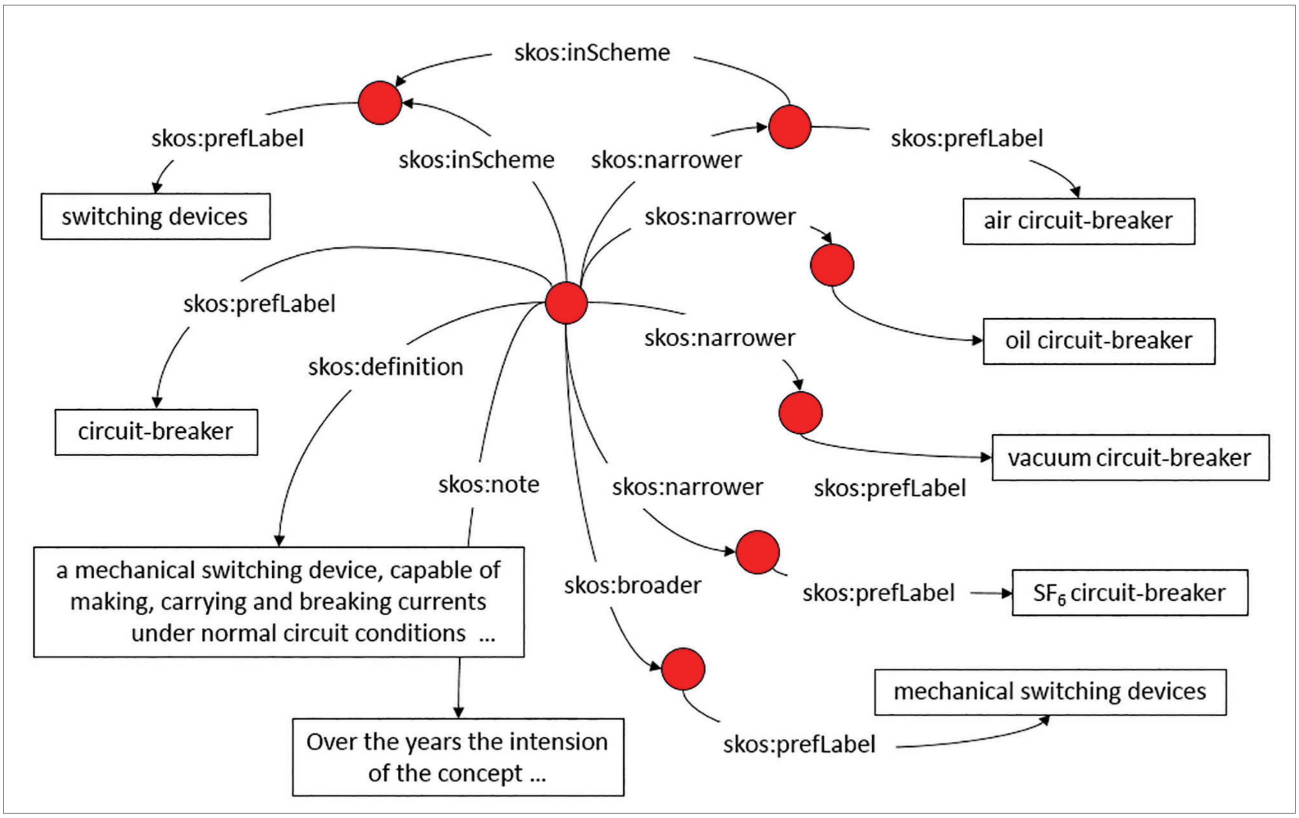

Figure 1: Graphical representation of a concept entry in SKOS

together within a dictionary entry. Onomasiological models (or concept-to-term models) on the other hand, consider senses, more precisely concepts, as the organizing principle of the terminology database, so that terminological information units are attached to the concepts that they refer to.

As previously stated, SKOS and TBX both provide concept-orientated data models. However, there remain considerable differences as far as the overall data organization, data granularity and styles are concerned. In the following sections, we focus on the characteristics of the two models, illustrating the analysis with examples using XML serialization. Other serializations are possible although this would not alter the underlying principle of what we want to demonstrate.

SKOS

SKOS was conceived to represent net-like multilingual knowledge organization systems, and thus to meet the demands of the Semantic Web and controlled vocabularies such as thesauri, classification schemes, taxonomies, or subject-heading systems. It is built upon RDF and RDFS and offers a very flat data model with basically only one formal level for storing data related to a specific concept.

In SKOS, a concept can be essentially represented and described through labels, semantic relationships, and documentary notes. More precisely, the following components form the core of a SKOS entry:

- <skos:Concept> is the single entry node to the entry structure;
- linguistic segments attached to the concept are represented by means of several possible relations: <skos:prefLabel>, <skos:altLabel>, and <skos:hiddenLabel>;
- additional documentary segments are attached to the concept with a variety of relations, such as <skos:definition>, <skos:note>, or <skos:example>;
- inter-concept relations allow the creation of a network of concepts, such as <skos:broader>, <skos:broaderTransitive>, <skos:narrower> or <skos:related>.

Figure 1 illustrates a graphical representation of a typical SKOS concept entry.

Figure 2 demonstrates an XML serialization of the sample SKOS represented by the graph in Figure 1, including two additional languages.

TBX

TBX, expressed by default in XML style, is the successor of the SGML-based terminology exchange format MARTIF

```
<rdf:RDF … xml:base ="http://www.www.tit-institute.de/">
  <skos:ConceptScheme rdf:about="switching_devices">
    <skos:prefLabel xml:lang="en">switching devices</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">appareils de connexion</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Schaltgeräte</skos:prefLabel>
  </skos:ConceptScheme>
  <skos:Concept rdf:about="switching_devices/concept4411420">
    <skos:prefLabel xml:lang="en">circuit-breaker</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">disjoncteur</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Leistungsschalter</skos:prefLabel>
    <skos:definition>a mechanical switching device, capable of making,
    carrying and breaking currents under normal circuit conditions and
    also making, carrying for a specified time and breaking currents under
    specified abnormal circuit conditions such as those of short circuit
    </skos:definition>
    <skos:note>Over the years the intension of the concept …</skos:note>
    <skos:broader rdf:resource="switching_devices/concept4411402"/>
    <skos:narrower rdf:resource="switching_devices/concept4411427"/>
    <skos:narrower rdf:resource="switching_devices/concept4411428"/>
    <skos:narrower rdf:resource="switching_devices/concept4411429"/>
    <skos:narrower rdf:resource="switching_devices/concept4411431"/>
    <skos:inScheme rdf:resource="switching_devices"/>
  </skos:Concept>
  <skos:Concept rdf:about="switching_devices/concept4411427">
    <skos:prefLabel xml:lang="en">air circuit-breaker</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">disjoncteur à air</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Luftleistungsschalter</skos:prefLabel>
    …
    <skos:inScheme rdf:resource="switching_devices"/>
  </skos:Concept>
<rdf:RDF>
```

Figure 2: Simple SKOS instance representing two concepts

[15] which both emerged from terminology management needs within the translation/localization (and even technical writing) industry. TBX was launched as a result of the EU project SALT [16], then further developed by LISA [17] and finally adopted to become ISO 30042 in 2008. TBX is based on the terminological metamodel, a strongly hierarchical data model described in ISO 16642 (see Figure 3), that allows for a very fine-grained description of terminological data.

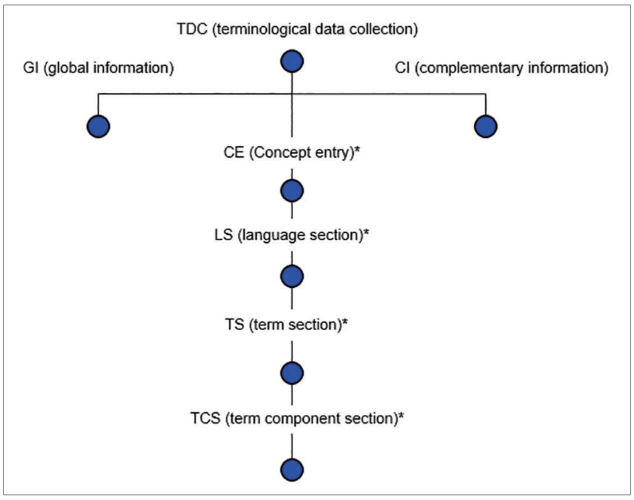

Figure 3: Terminological metamodel (adapted from ISO 16642:2017, 10)

As mentioned above, the metamodel reflects the principle of concept orientation, i. e. in a terminology database or terminological data collection (TDC) a concept entry (CE) describes a single concept. Starting from the top, the metamodel provides two high-level containers: a container for global information (GI) that applies to the complete terminology database (name of the terminology database, copyright information, history, etc.); and a container for complementary information (CI) such as complete bibliographical data sets or binary data, that is only documented once and then referenced from the respective concept entries.

Concept entries can be further structured by using additional language sections (LS) for instantiating languages and language-specific information. Each language section can include zero or more term sections (TS) for documenting terms and term-specific information, and individual words of a multiword term or morphemes of a single-word term can be described in the term component section (TCS).

Figure 4 illustrates a simplified version of the TBX core structure with the corresponding terminological metamodel levels in parentheses. For the sake of readability, the mandatory attributes for the TBX core structure elements have been omitted.

```
<tbxf> (= TDC)
  <tbxHeader>…</tbxHeader> (= GI)
  <text>
    <body>
      <conceptEntry> (= TE)
        <langSec> (= LS)
          <termSec> (= TS)
            <termCompSec>…</termCompSec> (= TCS)
          </termSec>
        </langSec>
      </conceptEntry>
    </body>
    <back>…</back> (= CI)
  </text>
</tbxf>
```

Figure 4: TBX core structure

In attribute style, almost all terminological data categories are instantiated as values of a type attribute that is associated with a core-structure element, such as <descrip>, <admin>, etc. (for example, <descrip type="subjectField"> $</descrip>). This mechanism provides sufficient flexibility for data exchange between different data models or terminology database definitions whilst ensuring a maximum of stringency for automated conversion.

The example in Figure 5 (on page 22) represents a concept entry in TBX for the concept circuit-breaker, including information relevant to the subject field, a definition and its source, references to the superordinate and various subordinate concepts, and the terms representing the concept in four languages, their respective sources and term-specific information such as status and part of speech.

Mapping the two worlds

The most relevant mapping scenarios between SKOS and TBX concern information related to the concept itself,

```
<conceptEntry id="teid_4411420">
    <descrip type="subjectField">switching devices</descrip>
    <descripGrp>
        <descrip type="definition">a mechanical switching device, capable of
        making, carrying and breaking currents under normal circuit conditions
        and also making, carrying for a specified time and breaking currents
        under specified abnormal circuit conditions such as those of short
        circuit</descrip>
        <admin type="sourceIdentifier" target="electropedia">Electropedia
        </admin>
    </descripGrp>
    <descrip type="superordinateConceptGeneric" target="teid_4411402">
    mechanical switching device</descrip>
    <descrip type="subordinateConceptGeneric" target="teid_4411427">air
    circuit-breaker</descrip>
    <descrip type="subordinateConceptGeneric" target="teid_4411428">oil
    circuit-breaker</descrip>
    <descrip type="subordinateConceptGeneric" target="teid_4411429">
    vacuum circuit-breaker</descrip>
    <descrip type="subordinateConceptGeneric" target="teid_4411431">SF6
    circuit-breaker</descrip>
    <langSec xml:lang="en">
        <termSec>
            <term>circuit-breaker</term>
            <termNote type="administrativeStatus">preferredTermAdmnSts
            </termNote>
            <termNote type="partOfSpeech">noun</termNote>
            <admin type="sourceIdentifier" target="electropedia">Electropedia
            </admin>
        </termSec>
    </langSec>
    <langSec xml:lang="fr">…</langSec>
    <langSec xml:lang="de">…</langSec>
</conceptEntry>
```

Figure 5: TBX document instance documenting one concept

concept relations, domain/subject field, definitions, and labels/terms.

Concepts

As we have seen, the two models are based on an onomasiological approach to the association between labels/terms and senses/concepts. As a result, they both organize their core informational content with the concept as the entry point. However, when looking at how both models organize their underlying conceptual systems, we can already identify a clear-cut difference in perspective.

The SKOS data model, taking its origins in the notion of thesaurus, puts the emphasis on the vertical (or hierarchical) organization of concepts as a KOS. By definition, a SKOS document instance describes an ontology that is articulated along a central generic-specific relation. In contrast, the data model implemented in TBX is more horizontal and based upon the idea of a conceptual space underlying a given subject field. The emphasis here is on the differentiation between concepts, which reflects the semantics of the terms used in this field. Consequently, even if <concept Entry> and <skos:Concept> are the obvious elements to be matched when mapping the two models, it is important to bear the difference in mind when comparing them more precisely.

Concept relations

In SKOS, the means for describing relationships between concepts is limited to three properties: broader, narrower and related. The SKOS Primer states that "skos:broader and skos:narrower enable the representation of hierarchical links, such as the relationship between one genre and its more specific species, or, depending on interpretations, the relationship between one whole and its parts" [see 18, 2.3]. In the traditional sense, this definition implies that the abstraction level between a genre and its species is equal to one. The examples given in the SKOS Primer underline this approach, as depicted in Figure 6:

```
ex:animals rdf:type skos:Concept;
    skos:prefLabel "animals"@en;
    skos:narrower ex:mammals

ex:mammals rdf:type skos:Concept;
    skos:prefLabel "mammals"@en;
    skos:broader ex:animals.
```

Figure 6: Broader and narrower relationship in SKOS

If abstraction levels greater than one for broader and narrower concepts come into play, SKOS proposes the additional properties <skos:broaderTransitive> and <skos:narrowerTransitive> (see Figure 7) which are generally destined to infer statements for reasoning purposes.

```
<skos:Concept rdf:about="switching_devices/concept4411401">
    <skos:prefLabel xml:lang="en">switching device</skos:prefLabel>
    <skos:definition>a device designed to make or break the current in one or
    more electric circuits</skos:definition>
    <skos:narrower rdf:resource="switching_devices/concept4411402"/>
    <skos:narrower rdf:resource="switching_devices/concept4411403"/>
    <skos:narrowerTransitive rdf:resource=
    "switching_devices/concept4411405"/>
    <skos:narrowerTransitive rdf:resource=
    "switching_devices/concept4411420"/>
    <skos:inScheme rdf:resource="switching_devices"/>
</skos:Concept>
<skos:Concept rdf:about="switching_devices/concept4411402">
    <skos:prefLabel xml:lang="en">mechanical switching device
    </skos:prefLabel>
    <skos:definition>a switching device designed to close and open one or
    more electric circuits by means of separable contacts</skos:definition>
    <skos:broader rdf:resource="switching_devices/concept4411401"/>
    <skos:narrower rdf:resource="switching_devices/concept4411405"/>
    <skos:narrower rdf:resource="switching_devices/concept4411420"/>
    <skos:inScheme rdf:resource="switching_devices"/>
</skos:Concept>
<skos:Concept rdf:about="switching_devices/concept4411405">
    <skos:prefLabel xml:lang="en">disconnector</skos:prefLabel>
    <skos:definition>a mechanical switching device which provides, in the
    open position, an isolating distance in accordance with specified
    requirements</skos:definition>
    <skos:broaderTransitive rdf:resource="switching_devices/concept4411401"
    />
    <skos:inScheme rdf:resource="switching_devices"/>
</skos:Concept>
```

Figure 7: Transitivity of concept relationships in SKOS

In contrast, TBX supports a wider variety of data categories that allow for the creation of comprehensive and fine-grained concept systems. As demonstrated in Figure 8, the relations between concepts can be specified by indicating their logical nature (generic, partitive, related), their relative position (superordinate, subordinate, coordinate, broader, narrower) or other kinds of properties (temporally, spatially or sequentially related). These properties are combined to make the relationships more explicit (/superordinateConceptGeneric/, /subordinateConceptPartitive/, /broaderConceptGeneric/, etc.).

```
<conceptEntry id="teid_4411402">
   <descrip type="superordinateConceptGeneric" target="teid_4411401">
   switching device</descrip>
   <descrip type="subordinateConceptGeneric" target="teid_4411420">
   circuit-breaker</descrip>
   <descrip type="narrowerConceptGeneric" target="teid_4411429">
   vacuum circuit-breaker</descrip>
   <langSec xml:lang="en">
      <termSec>
         <term>mechanical switching device</term>
      </termSec>
   </langSec>
</conceptEntry>
```

Figure 8: Concept relationships in TBX

Special attention should be drawn to the fact that TBX introduces a distinction with regard to hierarchically organized concepts depending on the number of abstraction levels that exist between different concepts. Data categories of the type /superordinate*/, for example, are defined in general terms (see 2.90 and 2.91) without specifying the number of abstraction levels, whereas data categories of the type /broader*/ are defined as concepts "two or more levels of abstraction higher than subject concept [sic.] in a generic hierarchical concept system" (see 2.10 and 2.11). Despite the lack of explicitness concerning the specification of the number of abstraction levels for /superordinate*/ and /subordinate*/, we consider that, in order for data category definitions to be consistent, the conclusion which has to be drawn from the /broader*/ definition is that /superordinate*/ and /subordinate*/ can only refer to an abstraction levels as being equal to one. In consequence, /broader*/ and /narrower*/, on the one hand, and <skos:broader> and <skos:narrower>, on the other hand, can be considered as false friends. Hence, <skos:broader> and <skos:narrower> should be mapped to /superordinate*/ and /subordinate*/ data categories; <skos:broaderTransitive> and <skos:narrowerTransitive> should be mapped to /broaderConcept*/ and /skos:narrowerConcept*/ respectively.

The one-to-one mapping of non-hierarchical relations can be easily achieved between <skos:related> (see 1.28) and /relatedConcept/ (see 2.70). However, if the entry structure of the terminology database requires further specification of the relation type, then the source database of the SKOS document instance or the SKOS document instance itself need to be adapted or transformed accordingly.

The <skos:*Match> element family is used for mapping between SKOS concept systems (see 1.4, 1.6, 1.12, 1.23, and 1.29) and therefore usually does not constitute a concern for the present SKOS <> TBX mapping analysis. However, if SKOS document instances contain several concept schemes, a preliminary analysis and determination of the preferred concept scheme or concepts should be made and the resulting prioritization should be implemented in the corresponding SKOS > TBX transformation stylesheet. In the reverse direction, <skos:*Match> does not need to be addressed, since an optimal terminology database does not include duplicates.

Domain or subject field

The difference is immediately reflected in the way concept systems are grounded in the two models. In TBX, concept entries are systematically associated with a subject field (<descrip type="subjectField">$</descrip>). The subject field, being a text description, does not imply any kind of structure except that terms are deemed unambiguous (monosemic) within a given field.

In SKOS, concepts are considered as autonomous units that may be linked to a concept scheme (<skos:ConceptScheme>) that encompasses all the concepts that are members of the same underlying KOS. In particular, there is, in theory, no constraint as to the width of the thematic coverage of a concept system. It may range from a simple group of concepts used to describe very local ontologies to very generic large-coverage ontologies [see 19] with no sense of term univocity or term prioritization within the corresponding field of knowledge. However, the SKOS designers recommend that no two concepts should have the same preferred lexical label in a given language when they belong to the same concept scheme [see 18, 2.2.1]. Yet, despite the recommendations given in these guidelines, it can be observed in large vocabularies such as IATE [20] or the UNESCO thesaurus [21] that these theoretical specificities tend to be dimmed down: subject fields and concept schemes appear to be used in quite similar ways in terms of expressing domains, and numerous terms are described under various concepts in both cases.

Therefore, when mapping between SKOS and TBX, one should be aware of the possible discrepancy between subject fields and concept schemes, even if, at first sight, they seem to express a similar reference system for the underlying concepts.

```
<rdf:RDF … xml:base ="http://www.www.tit-institute.de/">
  …
   <skos:inScheme rdf:resource="switching_devices"/>
  …
</rdf:RDF>

<descrip type="subjectField">switching devices</descrip>
```

Figure 9: Subject field instantiation in SKOS and TBX

Unlike TBX, SKOS provides an explicit access to entry points of subject field hierarchies using <skos:Concept

Scheme>, <skos:hasTopConcept>, <skos:TopConceptOf>, and <skos:inScheme> for hierarchization as shown in Figure 10.

```
<skos:ConceptScheme rdf:about="switching_devices">
    <skos:prefLabel xml:lang="en">switching devices</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">appareils de connexion</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Schaltgeräte</skos:prefLabel>
    <skos:hasTopConcept rdf:resource="switching_devices/concept4411402"/>
    <skos:hasTopConcept rdf:resource="switching_devices/concept4411403"/>
</skos:ConceptScheme>
<skos:Concept rdf:about="switching_devices/concept4411402">
    <skos:prefLabel xml:lang="en">mechanical switching device
    </skos:prefLabel>
    <skos:prefLabel xml:lang="fr">disjoncteur</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Leistungsschalter</skos:prefLabel>
    <skos:definition>a switching device designed to close and open one or
    more electric circuits by means of separable contacts</skos:definition>
    <skos:topConceptOf rdf:resource="switching_devices"/>
    <skos:inScheme rdf:resource="switching_devices"/>
</skos:Concept>
<skos:Concept rdf:about="switching_devices/concept4411403">
    <skos:prefLabel xml:lang="en">semiconductor switching device
    </skos:prefLabel>
    <skos:prefLabel xml:lang="fr">appareil de connexion à semiconducteur
    </skos:prefLabel>
    <skos:prefLabel xml:lang="de">Halbleiterschaltgerät</skos:prefLabel>
    <skos:definition>a switching device designed to make the current in an
    electric circuit by means of the controlled conductivity of a semiconductor
    </skos:definition>
    <skos:topConceptOf rdf:resource="switching_devices"/>
    <skos:inScheme rdf:resource="switching_devices"/>
</skos:Concept>
```

Figure 10: SKOS concept schemes

For larger databases covering several domains, SKOS offers <skos:collection> in conjunction with <skos:notation> and <skos:member> to build or merge these domain hierarchies into one single SKOS document instance as shown in Figure 11.

```
<skos:ConceptScheme rdf:about="thesaurus">
    ...
    <skos:hasTopConcept rdf:resource="thesaurus/concept4411201"/>
    <skos:hasTopConcept rdf:resource="thesaurus/concept4411301"/>
    <skos:hasTopConcept rdf:resource= thesaurus/concept4411401"/>
    ...
</skos:ConceptScheme>
<skos:Collection rdf:about="thesaurus/domain4">
    <skos:inScheme rdf:resource="thesaurus"/>
    <skos:notation>4</skos:notation>
    <skos:prefLabel xml:lang="en">electrical engineering</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">génie eléctrique</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Elektrotechnik</skos:prefLabel>
    ...
    <skos:member rdf:resource="thesaurus/mt4.41"/>
    <skos:member rdf:resource="thesaurus/mt4.42"/>
    ...
</skos:Collection>
<skos:Collection rdf:about="thesaurus/mt4.41">
    <skos:inScheme rdf:resource="thesaurus"/>
    <skos:notation>4.41</skos:notation>
    <skos:prefLabel xml:lang="en">switchgear, controlgear and fuses
    </skos:prefLabel>
    <skos:prefLabel xml:lang="fr">appareillage et fusibles</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Schaltanlagen und/oder Schaltgeräte und
     Sicherungen</skos:prefLabel>
    <skos:member rdf:resource="thesaurus/concept4411401"/>
    <skos:member rdf:resource="thesaurus/concept4411402"/>
    <skos:member rdf:resource="thesaurus/concept4411403"/>
    ...
</skos:Collection>
```

Figure 11: Domain classification in SKOS

TBX lacks explicit entry points to concept systems. They could only be inferred by (manually) searching for first level notations if available or by climbing up data categories pointing to preferred terms of superordinate concepts. Neither does TBX (yet) support sophisticated means for subject field classification. A workaround for mapping SKOS domain hierarchies to TBX could consist in using multiple “subject field” instances and specifying hierarchies with text content, as illustrated in Figure 12.

```
<termEntry id="teid_4411401">
    <descripGrp>
        <descrip type="subjectField">Domain 4: electrical engineering
        </descrip>
        <descrip type="subjectField">Main topic 4.41: switching devices
        </descrip>
    <descripGrp>
    ...
</termEntry>
```

Figure 12: Subject field classification in TBX

Definitions

<skos:definition> is always a child element of <skos:Concept> (see Figure 13). In TBX, however, definitions can be instantiated at the concept level or the language level (see Figure 14). This flexibility derives from the fact that terminology databases are usually designed for different use cases. In technical writing/authoring (in a monolingual

```
<skos:Concept rdf:about="switching_devices/concept4411420">
    <skos:prefLabel xml:lang="en">circuit-breaker</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">disjoncteur </skos:prefLabel>
    <skos:definition xml:lang="en">a mechanical switching device, capable of
    making, carrying and breaking currents under normal circuit conditions
    and also making, carrying for a specified time and breaking currents under
    specified abnormal circuit conditions such as those of short circuit
    </skos:definition>
    <skos:definition xml:lang="fr">appareil mécanique de connexion capable
    d'établir, de supporter et d'interrompre des courants dans les conditions
    normales du circuit, ainsi que d'établir, de supporter pendant une durée
    spécifiée et d'interrompre des courants dans des conditions anormales
    spécifiées du circuit telles que celles du court-circuit</skos:definition>
</skos:Concept>
```

Figure 13: Definitions in SKOS

```
<conceptEntry id="teid_4411420">
    <descripGrp>
        <descrip type="definition">a mechanical switching device, capable of
        making, carrying and breaking currents under normal circuit conditions
        and also making, carrying for a specified time and breaking currents
        under specified abnormal circuit conditions such as those of short
        circuit</descrip>
        <admin type="sourceIdentifier" target="electropedia">Electropedia
        </admin>
    </descripGrp>
    ...
</conceptEntry>

<conceptEntry id="teid_4411420">
    ...
    <langSec xml:lang="en">
        <descripGrp>
            <descrip type="definition">a mechanical switching device, capable
            of making, carrying and breaking currents under normal circuit
            conditions and also making, carrying for a specified time and
            breaking currents under specified abnormal circuit conditions such
            as those of short circuit</descrip>
            <admin type="sourceIdentifier" target="electropedia">Electropedia
            </admin>
        </descripGrp>
    </langSec>
    ...
</conceptEntry>
```

Figure 14: Definitions at concept and language level in TBX

environment), for instance, or in translation environments using a pivot language, definitions are frequently allocated to the concept level, whereas in translation environments without pivot languages definitions happen to appear at the language level in their respective languages.

As a consequence of the above, when mapping from TBX to SKOS, definitions are not problematic, since the instantiation of concept-related information in SKOS is restricted to a single level. In the reverse mapping scenario, however, the decision where to allocate a definition cannot be resolved without prior (human) disambiguation, even though there might be some indications that could lead to the assumption that the content of <skos:definition> could be mapped either to the concept or the language level in TBX. Let us suppose, for example, that there is one <skos:definition> per language, so we might assume that these definitions should be placed at the language level in TBX. Or imagine a SKOS instance with only one <skos:definition> per concept without any xml:lang attribute. We might conclude that these definitions should typically be mapped to the concept level in TBX. Nevertheless, it should not be forgotten that these considerations remain mere speculations as long as the TBX use case has not been previously specified.

Labels and terms

Intuitively, there should be a one-to-one mapping between the notions of labels and terms in the SKOS and TBX models respectively, since, with the broad onomasiological perspective, they both constitute linguistic representations that can be used to express the encompassing concept. However, when looking more closely at semantics in SKOS and TBX, some differences arise regarding the precise role they play.

Considering the label elements in the SKOS model from a general point of view, their role is focused on denotation, that is of being lexical expressions (single or multi word units) associated with the corresponding SKOS concept. They are also described along two single dimensions:

- the language of the label, expressed by means of the xml:lang attribute;
- the status of the label with regard to its relation to the concept, differentiating between preferred (<skos:pref Label>), alternative (<skos:altLabel>) and so-called hidden labels (<skos:hiddenLabel>).

In TBX, terms come with a more elaborate semantic set of descriptive features. The central idea that should be stressed here is term autonomy, which states that all the descriptive components of a term, comprising the term itself, should be grouped together with a dedicated reified object. Typically, this reification is reflected by the occurrence of a specific term section component in the terminological metamodel and the corresponding <termSec> element in TBX. Term autonomy allows terms to be described along a variety of possible dimensions:

- linguistic characteristics (part of speech, grammatical gender, etc.);
- discursive context (register, usage, etc.);
- term relations (abbreviation, equivalent, etc.);
- administrative characteristics (status, project, customer, source, etc.);
- management task (origination, modification, approval, etc.).

As can be seen, the only intersection of SKOS label properties and TBX term-specific data categories is the information related to the status of labels and terms.

<skos:prefLabel>, a property assigned to a resource for being the preferred label that represents the concept

(see 1.27), can be easily mapped to one of the two permissible TBX picklist values, that is “preferred” or “preferred term” (see Figure 15). The existence of more than one option here is due to the fact that TBX uses a more differentiated approach and classifies the term status information depending on whether the term pertains to a working environment (/administrativeStatus/, see 2.2) or to a normative environment (/normativeAuthorization/, see 2.54). Therefore, when transforming a SKOS document instance into TBX, this differentiation needs to be considered. The same applies to <skos:altLabel>, a property for alternative labels (see 5.2.1), and its natural counterparts, the two values “admitted” for /administrativeStatus/ and “admitted term” for /normativeAuthorization/. A transformation from TBX to SKOS presents a minor challenge in this regard as long as the working environment is not of importance.

```
<skos:prefLabel xml:lang="de">Vakuumleistungsschalter</skos:prefLabel>
<skos:altLabel xml:lang="de">Vakuumschalter</skos:altLabel>

<term>Vakuumleistungsschalter<term>
<termNote type="administrativeStatus">preferred</termNote>

<term>Vakuumschalter<term>
<termNote type="administrativeStatus">admitted</termNote>
```

Figure 15: Term status instantiation in SKOS and TBX

<skos:hiddenLabel> is to be used for labels that should only be accessible to applications for indexing or search operations, such as misspelled variants or other lexical labels, but which are not be visible otherwise (see 1.15). In this context, TBX provides some further specification apart from the working environment to which a term may pertain. In TBX, the data categories /administrativeStatus/ and /normativeAuthorirzation/ make it possible to instantiate the values “obsolete” and “superseded term” respectively, meaning terms that are “no longer preferred or admitted” – which matches part of the <skos:hiddenLabel> definition. However, the SKOS property “invisible” is not inherent in either TBX data category value definitions, and for good reason – obsolete or superseded terms may still need to be visible in authoring and translation environments as negative examples to be avoided during authoring or translation. Also, in some terminology databases, obsolete or superseded terms are maintained for documentary reasons. Similarly, the <skos:hiddenLabel> property “misspelled variants” has no semantic counterpart in TBX. As a matter of fact, misspelled variants are not usually included in terminology databases.

Nonetheless, if <skos:hiddenLabel> instantiates labels for indexing and search operations in the first place (including misspelled variants), /searchTerm/ (see 2.76) could be an alternative, as depicted in Figure 16.

```
<skos:hiddenLabel xml:lang="de">SF6 circuit breaker</skos:hiddenLabel>

<term>SF6 circuit-breaker<term>
<admin type="searchTerm">SF6 circuit breaker</admin>
```

Figure 16: Mapping misspelled variants from SKOS to /searchTerm/

Mapping from TBX to SKOS is less discretionary; preferred, admitted, and obsolete/superseded terms can be easily mapped to their corresponding counterparts <skos:prefLabel>, <skos:altLabel>, and <skos:hiddenLabel>. Other types of TBX term status information such as “legal”, “regulated” or “standardized” have no equivalent in SKOS.

SKOS instances may also include explicit relations between labels of one single concept, for instance, between a full form (sulphur hexafluoride circuit-breaker) and an abbreviated form of the label (SF6 circuit-breaker) or between a label and its translation. The first case is useful as long as more than two labels are provided for one concept in a given language; the second use case makes sense when the labels do not imply total synonymy, so that equivalencies between the labels in the corresponding languages need to be established. Figure 17 illustrates a link between two English labels by first defining the labels as first-order RDF resources which can then be linked by a <skosxl:labelRelation> statement.

```
<skosxl:Label rdf:about="SF6label1">
   <skosxl:literalForm xml:lang="en">SF6 circuit-breaker</skosxl:literalForm>
</skosxl:Label>
<skosxl:Label rdf:about="SF6label2">
   <skosxl:literalForm xml:lang="fr">disjoncteur à SF6</skosxl:literalForm>
</skosxl:Label>

<skosxl:altLabel rdf:about="SF6label1"/>
<skosxl:altLabel>
   <rdf:Description rdf:about="SF6label2">
      <skosxl:labelRelation rdf:resource="SF6label1"/>
   </rdf:Description>
</skosxl:altLabel>
```

Figure 17: Links between labels in different languages in SKOS

In TBX, relations between terms of a concept in different languages can be described using /transferComment/ (see Figure 18). However, /transferComment/ allows documenting different kinds of information (degree of equivalence, directionality, etc.; see 2.105), so, consequently, some further context information should be provided during transformation (such as “Equivalent French term:”) in order to refine the data.

```
<term>SF6 circuit-breaker<term>
<termNote type="transferComment">Equivalent French term: disjoncteur à
SF6</termNote>
```

Figure 18: Link between terms in different languages in TBX

As stated, the SKOS formalism illustrated in Figure 17 for linking terms in different languages can also be used to link an abbreviation with the corresponding full form of a label.

TBX provides specific data categories for these scenarios, as depicted in Figure 19.

```
<term>sulphur hexafluoride<term>
<termNote type="termType">full form</termNote>

<term>SF6<term>
<termNote type="abbreviatedFormFor">sulphur hexafluoride</termNote>
```

Figure 19: Link between the full form of a term and its abbreviation in TBX

Conclusion

As we have seen, the mapping rules we suggest only make sense if some constraints are implemented when designing one type of data source or the other – or, at least, when such constraints are not in place, that we can anticipate where information may be lost or represented with less felicity.

The major mapping problem lies in the difference between SKOS and TBX regarding the depth of the information description and the instantiation flexibility. TBX is a clear-cut, in-depth exchange format with high instantiation flexibility, whereas SKOS provides a relatively flat, rigid, and partially ambiguous description formalism. Consequently, the difference in the diversity of instantiable information units does not affect data integrity when mapping from SKOS to TBX, but generally leads to data impoverishment when mapping from TBX to SKOS.

Bibliography

[1] ISO 25964-1:2011 Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval. ISO: Geneva.

[2] ISO 25964-2:2013 Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies. ISO: Geneva.

[3] Park, Jack and Sam Hunting, Sam. 2002. XML Topic Maps: creating and using topic maps for the Web. Addison-Wesley Professional: Boston.

[4] Lei Zeng, Marcia, and Lois Mai Chan. 2004. “Trends and issues in establishing interoperability among knowledge organization systems.” Journal of the American Society for information science and technology 55.5: 377-395.

[5] World Wide Web Consortium (W3C). 2009a. “SKOS reference.” http://www.w3.org/TR/skos-reference.

[6] ISO 16642:2017 Computer applications in terminology – Terminological markup framework. ISO: Geneva.

[7] ISO 12620:2019 Management of terminology resources – Data category specifications. ISO: Geneva.

[8] ISO 26162:2019 Management of terminology resources – Terminology databases – Part 1: Design. Geneva: ISO.

[9] ISO 26162:2019 Management of terminology resources – Terminology databases – Part 2: Software. Geneva: ISO.

[10] ISO 30042:2019 Management of terminology resources – TermBase eXchange (TBX). ISO: Geneva.

[11] Reineke, Detlef. 2014. “TBX between termbases and ontologies.” In Terminology and Knowledge Engineering 2014 proceedings, ed. by TKE. https://hal.inria.fr/hal-00950862v2.

[12] World Wide Web Consortium (W3C). 2013. “OntoLex – Specification of Core Model.” https://www.w3.org/community/ontolex/wiki/Specification_of_Core_Model.

[13] World Wide Web Consortium (W3C). 2015. “Best Practices for Multilingual Linked Open Data community group, Guidelines for Linguistic Linked Data Generation: Multilingual Terminologies (TBX).” https://www.w3.org/2015/09/bpmlod-reports/multilingual-terminologies.

[14] Language Terminology/Translation and Acquisition Consortium (LTAC). 2019. “DatCatInfo.” http://www.datcatinfo.net.

[15] ISO 12200:1999 Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange. ISO: Geneva.

[16] Melby, Alan. 2015. “TBX: A terminology exchange format for the translation and localization industry.” In Handbook of Terminology, ed. by Hendrik J. Kockaert and Frieda Steurs, 393-424. Amsterdam/Philadelphia: John Benjamins.

[17] Romary, Laurent. 2001. “An abstract model for the representation of multilingual terminological data: TMF – Terminological Markup Framework.” In TAMA 2001 – Sharing Terminological Knowledge, Sharing Multilingual Content. Proceedings of the 5th TermNet Symposion, ed. by Infoterm. https://hal.inria.fr/inria-00100405.

[18] World Wide Web Consortium (W3C). 2009b. “SKOS Primer.” https://www.w3.org/TR/skos-primer.

[19] Hellmann, Sebastian, Jens Lehmann, and Sören Auer. 2009. “Learning of OWL class descriptions on very large knowledge bases.” International Journal on Semantic Web and Information Systems (IJSWIS) 5.2: 25-48.

[20] European Union (EU). 2019. “Interactive Terminology for Europe (IATE).” http://iate.europa.eu.

[21] UNESCO. 2019. “UNESCO Thesaurus.” http://vocabularies.unesco.org/browser/thesaurus.

Dr. Detlef Reineke is teaching and researching at the Faculty of Translation and Interpreting at the University of Las Palmas de Gran Canaria (Spain). He is currently chair of ISO/TC 37/SC 3 “Management of terminology resources”, expert in DIN, and also providing terminology consultancy and training for industry companies and government institutions.

Kontaktadresse
detlef.reineke@ulpgc.es

Laurent Romary is senior researcher at Inria (team ALMAnaCH, Paris, France). He is currently chair of ISO/TC 37 “Language and terminology” and involved in the development of various standards related to lexical data and language resources.

Kontaktadresse
laurent.romary@inria.fr